

# Algorithm that constructs two sequence-set betting strategies that predict all compressible sequences

January 1, 2013

## Abstract

Sequence-set betting is defined and an algorithm is described that constructs two sequence-set betting strategies such that on every non-Martin-Löf random sequence at least one of the strategies predicts the sequence.

## 1 Definition of sequence-set betting

A sequence-set betting strategy  $R$  can be defined as a pair of functions  $m^R$  and  $S^R$ . The mass function  $m^R$  is a function from words (finite binary sequences) to rationals and the sequence-set function  $S^R$  is a function from words to clopen sets of infinite binary sequences. For function  $m^R$  we require that  $m^R(v0) + m^R(v1) = m^R(v)$ . For function  $S^R$  we require that  $S^R(v0) \cup S^R(v1) = S^R(v)$  and that  $\lambda(S^R(v0)) = \lambda(S^R(v1)) = \frac{1}{2}\lambda(S^R(v))$  where  $\lambda$  is the Lebesgue measure of the set. A strategy  $R$  is total if  $m^R$  and  $S^R$  are defined for every word and it is computable if both functions are computable. The capital gained by strategy  $R$  on set of infinite sequences  $S^R(v)$  is  $m^R(v)/\lambda(S^R(v))$ . An infinite sequence  $\alpha$  is predicted by strategy  $R$  if there is an infinite set of words  $v_1, v_2, \dots$  such that  $\alpha \in S^R(v_i)$  and  $m^R(v_i)/\lambda(S^R(v_i)) > i$ . In other words  $\alpha$  is predicted by  $R$  if strategy has unbounded gain of capital while betting successively on  $\alpha$ . A sequence-set betting strategy  $R$  can be viewed as a set of triplets  $(v, S, m)$  where  $v$  is a word,  $S$  is a clopen set of infinite binary sequences and  $m$  is a rational. For a triplet  $t$  we denote the word with  $w(t)$ , the set of sequences with  $S(t)$  and the mass with  $m(t)$ .

## 2 Algorithm that constructs two sequence-set betting strategies

Theorem: There are two total and computable sequence-set betting strategies such that every non-Martin-Löf random sequence is predicted by at least one of

the strategies.

## 2.1 Algorithm overview

An algorithm constructs two sets of quintuples  $A$  and  $B$  that define two total and computable sequence-set betting strategies  $R^A$  and  $R^B$ . A quintuple  $q = (v, S, ms, me, L)$  consists of a word  $v$ , clopen set of sequences  $S$ , and rational values  $ms$ ,  $me$ ,  $L$ . We denote them with  $w(q)$ ,  $S(q)$ ,  $ms(q)$ ,  $me(q)$ ,  $L(q)$  respectively. We'll call the quintuples nodes. A triplet  $t$  of a strategy is defined by a node  $q$  as  $w(t) = w(q)$ ,  $S(t) = S(q)$ ,  $m(t) = ms(q) + me(q)$ . The algorithm iteratively adds nodes to both sets  $A$  and  $B$ . For each word  $v$  and any of the sets  $A$  or  $B$  there is some computation step at which a node  $q$  is added to the set such that  $w(q) = v$ . For nodes  $q, q'$  we say that  $q$  is a child node of  $q'$ ,  $q' \prec q$  if the word  $w(q')$  is a prefix of word  $w(q)$ ,  $w(q') \prec w(q)$ . A node in the set of nodes is a leaf node if there are no child nodes for that node in the set. For a set of nodes  $Q$  define the set of leaf nodes as  $LN(Q)$

$$LN(Q) = \{q : q \in Q, \forall q' \in Q, q \not\prec q'\}$$

Initially the algorithm adds the node  $a$  to set  $A$ ,  $A = \{a\}$  and node  $b$  to set  $B$ ,  $B = \{b\}$  such that the word of the node is an empty word  $\epsilon$ ,  $w(a) = w(b) = \epsilon$ . In the following computation steps the algorithm adds pairs of nodes  $q_0, q_1$  to one of the sets  $A, B$ . The nodes are such that if pair is to be added to set  $A$  there is a leaf node  $q$ ,  $q \in LN(A)$  (if the pair is to be added to  $B$  then there is  $q \in LN(B)$ ) and

$$\begin{aligned} w(q_0) &= w(q)0, \quad w(q_1) = w(q)1 \\ S(q) &= S(q_0) \cup S(q_1), \quad \lambda(S(q_0)) = \lambda(S(q_1)) \\ ms(q) &= ms(q_0) + ms(q_1), \quad me(q) = me(q_0) + me(q_1) \end{aligned}$$

The input for an algorithm instance are words  $v_a, v_b$  the basic interval  $w$  and the rationals  $m_a, m_b$ . An initial node  $a$  is added to  $A$  such that  $w(a) = v_a$ ,  $S(a) = w$ ,  $ms(a) + me(a) = m$  and an initial node  $b$  is added to  $B$  such that  $w(b) = v_b$ ,  $S(b) = w$ ,  $ms(b) + me(b) = m$ . We start the algorithm by running an instance of the algorithm with inputs  $v_a = v_b = \epsilon$ ,  $w = \{0|1\}^\infty$ ,  $m_a = m_b = 1$ . The algorithm instance calculates value  $k$  from  $(w, m_a, m_b)$  and enumerates a disjunct set of basic intervals  $P = \{p_1, p_2, \dots\}$ ,  $p_i \subset w$ , such that for each infinite sequence  $\alpha$  that has a  $k$ -compressible prefix and  $\alpha \in w$  there is some  $p_i$ ,  $\alpha \in p_i$  and  $\lambda(P) < 2^{-k+1}$ . In each iteration  $i$  of the algorithm instance,  $p_i$  from  $P$  is enumerated. Denote with  $P_n$  the union of so far enumerated intervals,  $P_n = \bigcup_{i=1}^n p_i$  where  $n$  is the number of iteration and  $P_0 = \{\}$ . Define  $LTS_n^A$  as the set of leaf nodes  $u$  in  $A$  at the beginning of iteration  $n$  such that  $a \preceq u$  and  $S(u) \cap P_{n-1} = \{\}$

$$LTS_n^A = \{u : u \in LN(A), a \preceq u, S(u) \cap P_{n-1} = \{\}\}$$

and define  $LTS_n^B$  as the set of leaf nodes  $v$  in  $B$  at the beginning of iteration  $n$  such that  $b \preceq v$  and  $S(v) \cap P_{n-1} = \{\}$

$$LTS_n^B = \{v : v \in LN(B), b \preceq v, S(v) \cap P_{n-1} = \{\}\}$$

For nodes  $u$ ,  $u \in LTS_n^A$  a finite set of child nodes is added to  $A$  and for nodes  $v$ ,  $v \in LTS_n^B$  a finite set of child nodes is added to  $B$  during iteration  $n$ . The added child nodes are such that for the added leaf nodes  $u'$  at the end of iteration  $n$  in  $A$ ,  $\{u' : u' \in LN(A), \exists u \in LTS_n^A, u \prec u'\}$ , we either have  $S(u') \cap p_n = \{\}$  or  $S(u') \subseteq p_n$ , and for the added leaf nodes  $v'$  in  $B$ ,  $\{v' : v' \in LN(B), \exists v \in LTS_n^B, v \prec v'\}$ , we either have  $S(v') \cap p_n = \{\}$  or  $S(v') \subseteq p_n$ . Furthermore for leaf node  $u'$  such that  $S(u') \subseteq p_n$  there is a leaf node  $v'$  in  $B$  such that  $S(v') = S(u') = w'$  where  $w'$  is a basic interval,  $w' \subseteq p_n$ . For these nodes we also have that  $ms(u') + me(u') > 0$ ,  $ms(v') + me(v') > 0$  and  $max(ms(u') + me(u'), ms(v') + me(v'))/\lambda(w') \geq 2max(m_a, m_b)/\lambda(w)$ . In other words at end of iteration  $n$  there will be some set of basic intervals  $\{w_1, \dots, w_z\}$  such that  $p_i = \bigcup_{j=1}^z w_j$  and for each  $w_j$  there will be two nodes  $u'$ ,  $v'$ , one in each strategy, such that  $S(u') = S(v') = w_j$  and at least one of the nodes doubles the initial capital of the algorithm instance. These node pairs are the initial nodes of another algorithm instances. The leaf nodes at the end of iteration  $n$  are the leaf nodes at the beginning of iteration  $n + 1$ . Note that we have  $\bigcup_{u \in LTS_{n+1}^A} S(u) = \bigcup_{v \in LTS_{n+1}^B} S(v) = w \setminus P_n$ . It remains to show that algorithm instance can calculate  $k$  from the inputs and that it can at least double the capital on the infinite sequences in the set  $P$ .

## 2.2 Initialization step

The input for an algorithm instance are words  $v_a$ ,  $v_b$  the basic interval  $w$  and the rationals  $m_a$ ,  $m_b$ . Set the number of iteration  $n = 0$ . Define

$$m = \min(m_a, m_b)/2$$

Add initial node  $a = (v_a, w, m, m_a - m, 0)$  to the set of nodes  $A$ . Add initial node  $b = (v_b, w, m, m_b - m, 0)$  to the set of nodes  $B$ . We have that  $ms(a) = ms(b) = m$ , the  $ms$  is the part of the mass used by the algorithm instance to increase the capital on infinite sequences in  $P$  and the remainder  $me$  is the reserved part of mass that ensures that all nodes have mass greater than 0. The  $me$  part of mass is distributed evenly among the child nodes. In all iterations  $n$  for nodes  $a' \in LTS_n^A$ ,  $a \prec a'$  we have  $me(a') = \frac{\lambda(S(a'))}{\lambda(w)} me(a)$  and analogously we have for nodes  $b' \in LTS_n^B$ ,  $b \prec b'$   $me(b') = \frac{\lambda(S(b'))}{\lambda(w)} me(b)$ . The capital gained on the infinite sequences in  $P$  will be at least  $c = 2(m_a + m_b)/\lambda(w)$  for at least one of the strategies  $R^A$ ,  $R^B$ . We choose the value  $k$  such that the measure of the enumerated set  $P$  is

$$\lambda(P) < \frac{m^2(1 - cs)}{8\lambda(w)c^2(1 + cs)^2} \quad (1)$$

Where  $cs$  is some predetermined constant  $0 < cs < 1$ . At the beginning of iteration  $n$  for leaf nodes  $a' \in LTS_n^A, b' \in LTS_n^B$  such that  $S(a') \cap S(b') \neq \{\}$  we will have

$$\lambda(w)\lambda(S(a') \cap S(b')) < (1 + cs)(\lambda(S(a')) + L(a'))(\lambda(S(b')) + L(b')) \quad (2)$$

$$\lambda(w)\lambda(S(a') \cap S(b')) > (1 - cs)(\lambda(S(a')) + L(a'))(\lambda(S(b')) + L(b')) \quad (3)$$

In particular, at the beginning of first iteration,  $n = 1$ , we have only one leaf node  $a'$  in  $A$  such that  $a \preceq a'$ , the node  $a$  itself,  $a' = a$  and in  $B$  we have only one node  $b'$ , such that  $b \preceq b'$ , the node  $b$  itself,  $b' = b$  and we have  $\lambda(w)\lambda(S(a') \cap S(b')) = (\lambda(S(a')) + L(a'))(\lambda(S(b')) + L(b'))$ .

### 2.3 Iteration of the algorithm

Increment the number of iteration  $n := n + 1$ . Enumerate the next basic interval  $p_n \in P$ . Next we add child nodes to all  $u \in LTS_n^A$  such that there is  $v \in LTS_n^B$  such that  $S(u) \cap S(v) \cap p_n \neq \{\}$  and  $S(u) \cap S(v) \not\subseteq p_n$ . Define  $LTN_n^A$  as the set of leaf nodes in  $A$  such that  $a \preceq u$  and  $S(u) \cap P_{n-1} = \{\}$  after these child nodes are added in iteration  $n$ .

$$LTN_n^A = \{u : u \in LN(A), a \preceq u, S(u) \cap P_{n-1} = \{\}\}$$

The leaf nodes  $u \in LTN_n^A, v \in LTS_n^B$  for which we have  $S(u) \cap S(v) \cap p_n \neq \{\}$  we also have  $S(u) \cap S(v) \subseteq p_n$ . For  $u \in LTN_n^A, u' \in LTS_n^A, u' \preceq u$  we have

$$ms(u) = \frac{\lambda(S(u))}{\lambda(S(u'))} ms(u') \quad (4)$$

$$L(u) = \frac{\lambda(S(u))}{\lambda(S(u'))} L(u') \quad (5)$$

The clopen sets  $S(u)$  for  $u \in LTN_n^A$  are such that the (2) and (3) are valid for nodes  $u \in LTN_n^A, v \in LTS_n^B, S(u) \cap S(v) \neq \{\}$ . The construction of sets of sequences for nodes in  $LTN_n^A$  is described in 2.5.1. Next we iterate in any order through pairs of leaf nodes  $(u, v), u \in LTN_n^A, v \in LTS_n^B$  for which we have  $S(u) \cap S(v) \subseteq p_n$  and determine which of the strategies will gain capital of at least  $c$  on infinite sequences in the intersection  $S(u) \cap S(v)$ . Denote with  $d^A(u, v)$  the part of mass  $ms(u)$  assigned to the intersection  $S(u) \cap S(v)$  by strategy  $A$  and with  $d^B(u, v)$  the part of mass  $ms(v)$  assigned to the intersection by strategy  $B$ . Initially set  $d^A(u, v) = d^B(u, v) = 0$ . Denote with  $ms'(u) = ms(u) - \sum_{v \in LTS_n^B} d^A(u, v)$ , initially we have  $ms'(u) = ms(u)$ . Analogously for  $B$  denote  $ms'(v) = ms(v) - \sum_{u \in LTN_n^A} d^B(u, v)$ , initially  $ms'(v) = ms(v)$ .

**Mass assignment step (MAS):** Pick some leaf node pair  $(u, v)$ ,  $u \in LTN_n^A$ ,  $v \in LTS_n^A$  for which we have  $S(u) \cap S(v) \subseteq p_n$  and  $d^A(u, v) = d^B(u, v) = 0$ . If

$$(ms'(u) - 2c\lambda(S(u) \cap S(v)))/(S(u) + L(u)) \geq ms'(v)/(\lambda(S(v)) + L(v))$$

then set

$$d^B(u, v) = 0, d^A(u, v) = 2c\lambda(S(u) \cap S(v))$$

If

$$ms'(u)/(S(u) + L(u)) \leq (ms'(v) - 2c\lambda(S(u) \cap S(v)))/(\lambda(S(v)) + L(v))$$

then set

$$d^A(u, v) = 0, d^B(u, v) = 2c\lambda(S(u) \cap S(v))$$

Otherwise find  $da, db$  such that

$$da + db = 2c\lambda(S(u) \cap S(v))$$

and

$$(ms'(u) - da)/(S(u) + L(u)) = (ms'(v) - db)/(\lambda(S(v)) + L(v))$$

then set

$$d^A(u, v) = da, d^B(u, v) = db$$

If we haven't iterated through all leaf node pairs for which  $S(u) \cap S(v) \subseteq p_n$  goto MAS.

For the leaf nodes  $u' \in LTN_n^A$  and  $v' \in LTS_n^B$  we add child nodes to  $A$  and  $B$ . The child nodes added to  $A$  are such that for the leaf nodes we have

$$u \in LN(A), \exists u' \in LTN_n^A, u' \prec u \Rightarrow S(u) \cap p_n = \{\} \vee S(u) \subseteq p_n \quad (6)$$

Analogously for child nodes added to  $B$  we have that the leaf nodes  $v \in LN(B)$ ,  $\exists v' \in LTS_n^B$ ,  $v' \prec v$  are such that either  $S(v) \cap p_n = \{\}$  or  $S(v) \subseteq p_n$ . For the leaf nodes such that  $S(u) \cap p_n = \{\}$  we set

$$ms(u) = ms'(u')\lambda(S(u))/\lambda(S(u') \setminus p_n) \quad (7)$$

$$L(u) = (L(u') + \lambda(S(u') \cap p_n))\lambda(S(u))/\lambda(S(u') \setminus p_n) \quad (8)$$

and analogously for the leaf nodes in  $B$  such that  $S(v) \cap p_n = \{\}$ . The sets  $S(u)$  and  $S(v)$  are such that the (2) and (3) are valid if  $S(u) \cap S(v) \neq \{\}$ . The construction of sets of sequences for the child nodes that are added to  $A$  and  $B$  is described in 2.5.2.

For the leaf nodes such that  $S(u) \subseteq p_n$  there is a leaf node  $v$  in  $B$   $v \in LN(B)$ ,  $\exists v' \in LTS_n^A$ ,  $v' \prec v$  such that  $S(v) = S(u) = w'$  where  $w'$  is some basic interval that contains sequences from  $p_n$ ,  $w' \subseteq p_n$ . The nodes

$u, v$  are the nodes added in the initialization step of a new algorithm instance. If the sequence set of a node is a subset of enumerated interval  $p_n$ , some mass might be left unassigned after MASs. Denote the unassigned mass

of a node  $q$  with  $um(q)$ ,  $um(q) = \begin{cases} ms'(q) & \text{if } S(q) \subseteq p_n \\ 0 & \text{otherwise} \end{cases}$ . We set  $m_a =$

$\frac{\lambda(w')}{\lambda(w)}me(a) + d^A(u', v')\lambda(S(u))/\lambda(S(u') \cap S(v')) + um(u')\lambda(S(u))/\lambda(S(u'))$  and  $m_b = \frac{\lambda(w')}{\lambda(w)}me(b) + d^B(u', v')\lambda(S(v))/\lambda(S(u') \cap S(v')) + um(v')\lambda(S(v))/\lambda(S(v'))$  and we start another instance of the algorithm for inputs  $(u, v, w', m_a, m_b)$ .

If after some MAS for pair  $a', b'$  the values  $ms'(a')$ ,  $ms'(b')$  become non-positive then the algorithm fails to assign enough mass to the intersection in order to at least double the capital on the infinite sequences in the intersection. If we can prove that this never happens we have proven the theorem.

## 2.4 Proof

Define a function

$$indi(v) = \begin{cases} i & \text{if } v \in LTS_i^B \\ \text{undefined} & \text{otherwise} \end{cases}$$

Denote the leaf node pair  $(a', b')$  picked in  $j$ th MAS of iteration  $i$  with  $lnp(i, j)$ ,  $lnp(i, j) = (a', b')$ . For some iteration  $i$ , node  $v \in LTS_i^B$  and MAS  $j$  denote with  $E_{v,j}$  the set of leaf-nodes from strategy  $A$  that have been picked up to and including step  $j$ .

$$E_{v,j} = \{u : \exists e \in [1, j], lnp(indi(v), e) = (u, v)\}$$

Denote the remaining mass for node  $a' \in LTN_i^A$  after MAS  $j$  with  $ms_j(a')$ . If  $ms_j(a') < ms_{j-1}(a')$  then there is some  $b' \in LTS_i^B$ ,  $lnp(i, j) = (a', b')$  such that  $S(a') \cap S(b') \subseteq p_i$  and  $d^A(a', b') = ms_{j-1}(a') - ms_j(a')$ . Denote with  $f$  the value

$$f = \frac{\lambda(w)}{m}ms_{j-1}(a')/(\lambda(S(a')) + L(a'))$$

and with  $f'$  the value

$$f' = \frac{\lambda(w)}{m}ms_j(a')/(\lambda(S(a')) + L(a'))$$

Denote  $const = \frac{m^2(1-cs)}{4\lambda(w)c^2(1+cs)^2}$ . We show in 2.5.3 that

$$\lambda(S(a') \cap S(b')) + L(b') + \sum_{u \in E_{b',j-1}} \lambda(S(u) \cap S(b')) > const(1 - f')(f - f') \quad (9)$$

Assume that in some iteration  $r$  and after some MAS  $t$  for some leaf nodes  $a' \in LTN_r^A$   $ms_t(a') \leq 0$ ,  $ms_{t-1}(a') > 0$ . Denote with  $F$  the set of node pairs

$(u, v)$  such that in some iteration  $i$  and MAS  $j$  the value  $d^A(u, v) > 0$  and  $u \preceq a'$  and  $i \leq r$ .

$$F = \{(u, v) : \exists i, j, i \leq r, \text{ lnp}(i, j) = (u, v), u \preceq a', d^A(u, v) > 0\}$$

Define a function

$$\text{indj}(u, v) = \begin{cases} j & \exists i, j \text{ lnp}(i, j) = (u, v) \\ \text{undefined} & \text{otherwise} \end{cases}$$

We show in 2.5.4 that

$$\sum_{(u, v) \in F} (\lambda(S(u) \cap S(v)) + L(v) + \sum_{u' \in E_v, \text{indj}(u, v) = 1} \lambda(S(u') \cap S(v))) \leq \lambda(P_r)$$

We show in 2.5.5 that by summing the expression (9) over the node pairs in  $F$  we have

$$\sum_{(u, v) \in F} (\lambda(S(u) \cap S(v)) + L(v) + \sum_{u' \in E_v, \text{indj}(u, v) = 1} \lambda(S(u') \cap S(v))) > \text{const}/2$$

It follows that  $\lambda(P_r) \geq \text{const}/2$  which is in contradiction with (1) from which follows that in all iterations and MASs and for all leaf nodes  $a', b'$   $ms'(a') > 0$ ,  $ms'(b') > 0$ .

## 2.5 Proofs for used statements

### 2.5.1

The leaf node  $a' \in LTS_n^A$  has intersection with some set of leaf nodes  $\{b_1, \dots, b_z\}$ ,  $b_i \in LTS_n^B$ ,  $S(a') \cap S(b_i) \neq \{\}$  and  $S(a') = \bigcup_{i=1}^z S(a') \cap S(b_i)$ . We have that for all  $i \in [1, \dots, z]$  (2) and (3) are valid for node pair  $a', b_i$ . If for some  $i$  the intersection  $S(a') \cap S(b_i)$  contains both the sequences in the enumerated basic interval  $p_n$  and the sequences which are not in  $p_n$ , we add child nodes to node  $a'$ . The added child nodes are such that there are  $2^m$  leaf nodes  $u$ ,  $a' \prec u$ . The measure of the leaf nodes sequence-set is  $\lambda(S(u)) = \lambda(a')2^{-m}$ . The value  $L(a')$  is evenly distributed among leaf nodes,  $L(u) = 2^{-m}L(a')$ . For all  $i \in [1, \dots, z]$  and all the added leaf nodes  $u$  the intersection  $S(u) \cap S(b_i)$  contains either the sequences in the  $p_n$  or the sequences which are not in  $p_n$  and the node pair  $u, b_i$  satisfies the equations (2) and (3).

In order to determine the sequence-set of the leaf nodes  $u$  we construct  $z2^m$  disjunct, finite sets of basic intervals  $VS_j^i$  where  $i \in [1, \dots, z]$  and  $j \in [1, \dots, 2^m]$ . Denote the union of basic intervals  $e$  in  $VS_j^i$  with  $UVS_j^i = \bigcup_{e \in VS_j^i} e$ . If we index

the added leaf nodes  $u_1$  through  $u_{2^m}$  then  $UVS_j^i$  will correspond to intersection  $S(u_j) \cap S(b_i)$ . The requirements for  $VS_j^i$  are:

$$\text{(Req1)} \quad \forall i, i \in [1, \dots, z], \bigcup_{j=1}^{2^m} UVS_j^i = S(a') \cap S(b_i)$$

$$\text{(Req2)} \quad \forall j, j \in [1, \dots, 2^m], \lambda\left(\bigcup_{i=1}^z UVS_j^i\right) = 2^{-m} \lambda(S(a'))$$

$$\text{(Req3)} \quad \forall i, i \in [1, \dots, z], \forall j, j \in [1, \dots, 2^m]$$

$$((\forall e \in VS_j^i \ e \subseteq p_n) \vee (\forall e \in VS_j^i \ e \cap p_n = \{\}))$$

$$\text{(Req4)} \quad \forall i, i \in [1, \dots, z], \forall j, j \in [1, \dots, 2^m]$$

$$1 - cs < \frac{\lambda(w)\lambda(UVS_j^i)}{2^{-m}(\lambda(S(a')) + L(a'))(\lambda(S(b_i)) + L(b_i))} < 1 + cs$$

From (Req1) we have that every sequence from the sequence-set of node  $a'$  is contained in some basic interval  $e$  of some set  $VS_j^i$  since  $S(a') = \bigcup_{i=1}^z \bigcup_{j=1}^{2^m} UVS_j^i$ .

From (Req2) we have that the measure of the sequence-set for the added leaf nodes  $u$  is  $\lambda(S(u)) = \lambda(a')2^{-m}$  since  $S(u_j) = \bigcup_{i=1}^z UVS_j^i$ . From (Req3) we have that for all the added leaf nodes  $u_j$  the intersection  $S(u_j) \cap S(b_i)$  contains either the sequences in the  $p_n$  or the sequences which are not in  $p_n$  since  $S(u_j) \cap S(b_i) = UVS_j^i$ . The (Req4) shows that the equations (2) and (3) are valid for node pairs  $u_j, b_i$  since  $\lambda(S(u_j)) + L(u_j) = 2^{-m}(\lambda(S(a')) + L(a'))$ .

Next we construct the sets of basic intervals  $VS_j^i$ . Denote with

$$t_i = \frac{\lambda(w)\lambda(S(a') \cap S(b_i))}{(\lambda(S(a')) + L(a'))(\lambda(S(b_i)) + L(b_i))}$$

We have from (2) and (3) that  $1 - cs < t_i < 1 + cs$ .

Find  $h$  such that  $1 - cs < t_i - 2^{-h+1}$  and  $t_i + 2^{-h+1} < 1 + cs$  for all  $i \in [1, z]$ .

Since  $S(a')$ ,  $S(b_i)$  and  $p_n$  are clopen sets there is some  $d$  such that  $S(a') = \bigcup_{j=1}^{2^d} v_j$  where  $v_j$  is a basic interval such that  $\lambda(v_j) = 2^{-l(w(a'))-d}$  and  $v_j \subseteq S(a') \cap S(b_i)$  for some  $i \in [1, z]$  and either  $v_j \subseteq p_n$  or  $v_j \cap p_n = \{\}$ .

Set  $m = 4d + h + 1$  where  $2^m$  is the number of the added leaf nodes.

Denote the set of basic intervals  $v$  of measure  $\lambda(v) = 2^{-l(w(a'))-d}$  which are contained in sequence-set of node  $a'$  with  $V$

$$V = \{v : \lambda(v) = 2^{-l(w(a'))-d}, v \subseteq S(a')\}$$

We have that the number of intervals in  $V$  is  $2^d$ . Denote with  $r_i$  the number of basic intervals in  $V$  which are subsets of the enumerated interval  $p_n$  and are subsets of the intersection  $S(a') \cap S(b_i)$

$$r_i = |\{v : v \in V, v \subseteq S(a') \cap S(b_i), v \subseteq p_n\}|$$



and with  $g_i$  denote the number of basic intervals in  $V$  which don't contain sequences from  $p_n$  and are subsets of the intersection  $S(a') \cap S(b_i)$

$$g_i = |\{v : v \in V, v \subseteq S(a') \cap S(b_i), v \cap p_n = \{\}\}|$$

Then  $\sum_{i=1}^z (r_i + g_i) = 2^d$ . Denote the set of basic intervals  $v$  of measure  $\lambda(v) = 2^{-l(w(a'))-d-m-h}$  which are contained in sequence-set of node  $a'$  with  $VS$

$$VS = \{vs : vs \subset S(a'), \lambda(vs) = 2^{-l(w(a'))-d-m-h}\}$$

These are the intervals that will be distributed among the sets  $VS_j^i$ . Denote the intervals from  $VS$  that are contained in intersection  $S(a') \cap S(b_i) \cap p_n$  with  $RS_i$

$$RS_i = \{vs : vs \in VS, vs \subset S(a') \cap S(b_i), vs \subset p_n\}$$

and the intervals from  $VS$  that are contained in intersection  $S(a') \cap S(b_i)$  but are not contained in  $p_n$  with  $GS_i$

$$GS_i = \{vs : vs \in VS, vs \subset S(a') \cap S(b_i), vs \cap p_n = \{\}\}$$

We have that  $|RS_i| = r_i 2^{m+h}$  and  $|GS_i| = g_i 2^{m+h}$ . Let's say that for the first  $x$  nodes in set  $\{b_1, \dots, b_z\}$  the intersection  $S(a') \cap S(b_i)$  contains both the sequences from the interval  $p_n$  and the sequences not in  $p_n$ , that is both  $RS_i \neq \{\}$  and  $GS_i \neq \{\}$  for  $i \in [1, x]$ . For the remaining nodes the intersection  $S(a') \cap S(b_i)$  contains either the sequences in  $p_n$  or the sequences not in  $p_n$ , that is either  $RS_i = \{\}$  or  $GS_i = \{\}$  for  $i \in [x+1, z]$ .

Denote  $f_i = \lfloor \frac{2^m}{r_i + g_i} \rfloor$  and  $o_i = 2^m - f_i(r_i + g_i)$ .

Define a set of indices  $\{n_0, n_1, \dots, n_x\}$ . Set  $n_0 = 0$  and for the rest set  $n_i = n_{i-1} + 2o_i r_i 2^h$ . For a given  $i \in [1, z]$  the  $VS_j^i$  will contain the same number of elements,  $|VS_j^i| = 2^h(r_i + g_i)$ , for most  $j \in [1, 2^m]$  except for the  $j$  defined by the indices  $\{n_0, n_1, \dots, n_x\}$ . For  $i \in [2, x]$  the  $VS_j^i \neq 2^h(r_i + g_i)$  only for  $j \in [n_{i-2} + 1, n_i]$ . For  $i = 1$  the  $j$  on which the number of elements in  $VS_j^1$  is different from  $2^h(r_1 + g_1)$  depends on whether there are some leaf nodes  $b_i$  for which the intersection  $S(a') \cap S(b_i)$  contains only sequences in  $p_n$  or only sequences not in  $p_n$ . If there are such nodes then  $x < z$ , otherwise  $x = z$ .

In case  $x = z$ , for  $i = 1$  the  $VS_j^1 \neq 2^h(r_1 + g_1)$  only for  $j \in [1, n_1] \cup [n_{x-1} + 1, n_x]$ .

In case  $x < z$  the  $VS_j^1 \neq 2^h(r_1 + g_1)$  only for  $j \in [1, n_1]$  and the  $VS_j^{x+1} \neq 2^h(r_{x+1} + g_{x+1})$  only for  $j \in [n_{x-1} + 1, n_x]$ . For  $i \in [x+2, z]$  we will have  $|VS_j^i| = 2^h(r_i + g_i)$  for all  $j \in [1, 2^m]$ .

Define function  $next(i) = \begin{cases} 1 & \text{if } i = z \\ i + 1 & \text{otherwise} \end{cases}$ . Denote  $i' = next(i)$ .

Define  $PS_i = \begin{cases} RS_i & \text{if } RS_i \neq \{\} \\ GS_i & \text{otherwise} \end{cases}$ . We have that for  $i \in [1, x]$  the  $PS_i$  is

always  $RS_i$  since both  $RS_i$  and  $GS_i$  are nonempty. For  $i \in [x+1, z]$  either  $RS_i$  or  $GS_i$  are empty and  $PS_i$  is the nonempty of the two sets.

For each  $i \in [1, x]$  for the  $j$  in the first half of the interval  $[n_{i-1} + 1, n_i]$ , that is the interval  $[n_{i-1} + 1, n_i - o_i r_i 2^h]$ , we put  $2^h(r_i + g_i) + 1$  elements from  $RS_i$  into  $VS_j^i$  and in the second half of the interval  $[n_{i-1} + 1, n_i]$ , that is the interval  $[n_i - o_i r_i 2^h + 1, n_i]$ , we put  $2^h(r_i + g_i) - 1$  elements from  $GS_i$  into  $VS_j^i$ . At the same time for the  $j$  in the first half of the interval  $[n_{i-1} + 1, n_i]$  we put  $2^h(r_{i'} + g_{i'}) - 1$  elements from  $PS_{i'}$  into  $VS_j^{i'}$ , and in the second half of that interval we put  $2^h(r_{i'} + g_{i'}) + 1$  elements from  $PS_{i'}$  into  $VS_j^{i'}$ . In the remaining  $VS_j^i$  we put  $2^h(r_i + g_i)$  elements from  $VS$  and all of these elements are either from  $GS_i$  or from  $RS_i$ .

The construction of  $VS_j^i$  is:

- For  $i \in [1, x]$  for  $j \in [n_{i-1} + 1, n_i - o_i r_i 2^h]$  we have  $|VS_j^i| = 2^h(r_i + g_i) + 1$ ,  $VS_j^i \subset RS_i$  and  $|VS_j^{i'}| = 2^h(r_{i'} + g_{i'}) - 1$ ,  $VS_j^{i'} \subset PS_{i'}$
- For  $i \in [1, x]$  for  $j \in [n_i - o_i r_i 2^h + 1, n_i]$  we have  $|VS_j^i| = 2^h(r_i + g_i) - 1$ ,  $VS_j^i \subset GS_i$  and  $|VS_j^{i'}| = 2^h(r_{i'} + g_{i'}) + 1$ ,  $VS_j^{i'} \subset PS_{i'}$
- For all the remaining  $VS_j^i$  we have  $|VS_j^i| = 2^h(r_i + g_i)$  and either  $VS_j^i \subset GS_i$  or  $VS_j^i \subset RS_i$

We first verify (Req1) for  $i \in [x + 2, z]$ . We have that for each  $j \in [1, 2^m]$  the  $VS_j^i$  contains  $2^h(r_i + g_i)$  elements from  $PS_i$ , and therefore  $\sum_{j \in [1, 2^m]} |VS_j^i| = 2^{m+h}(r_i + g_i)$ . Since for  $i \in [x + 1, z]$  the  $PS_i$  contains  $2^{m+h}(r_i + g_i)$  elements and  $\bigcup_{vs \in PS_i} vs = S(a') \cap S(b_i)$  we have that the (Req1) is satisfied for  $i \in [x + 2, z]$ .

For  $i = x + 1$  we have that for  $j$  in the first half of interval  $[n_{x-1} + 1, n_x]$  the  $VS_j^{x+1}$  contains  $2^h(r_{x+1} + g_{x+1}) - 1$  elements from  $PS_i$ , for  $j$  in the second half of interval  $[n_{x-1}, n_x]$  the  $VS_j^{x+1}$  contains  $2^h(r_{x+1} + g_{x+1}) + 1$  elements from  $PS_i$  and for all the other  $j$  the  $VS_j^{x+1}$  contains  $2^h(r_{x+1} + g_{x+1})$  elements from  $PS_i$ . Therefore  $\sum_{j \in [1, 2^m]} |VS_j^{x+1}| = 2^{m+h}(r_{x+1} + g_{x+1})$  and the (Req1) is satisfied for  $i = x + 1$ . We also have to show that that  $n_x \leq 2^m$ . We have  $2^m = 2^{h+1}(\sum_{i=1}^z (r_i + g_i))^4$ . We have  $n_x = \sum_{i \in [1, x]} 2o_i r_i 2^h$  and since  $o_i < r_i + g_i$  we have  $n_x < 2^{h+1} \sum_{i \in [1, x]} (r_i + g_i)^2$ .

We now verify (Req1) for  $i \in [1, x]$ . From  $f_i = \frac{2^m - o_i}{r_i + g_i}$  we have that  $|RS_i| = r_i 2^{m+h} = f_i r_i 2^h(r_i + g_i) + o_i r_i 2^h$  and  $|GS_i| = g_i 2^{m+h} = (f_i g_i + o_i) 2^h(r_i + g_i) - o_i r_i 2^h$ . Note that  $f_i r_i + f_i g_i + o_i = 2^m$ .

There are  $f_i r_i$  sets  $VS_j^i$  such that  $VS_j^i \subset RS_i$ . Of those  $f_i r_i$  sets there are  $o_i r_i 2^h$  sets that contain  $2^h(r_i + g_i) + 1$  elements from  $RS_i$ , these are the sets  $VS_j^i$  where  $j$  is in the first half of the interval  $[n_{i-1} + 1, n_i]$ . The remainder of the sets contain on average  $2^h(r_i + g_i)$  elements. That is if  $i = 1$  and  $z > x$  each of the remaining  $f_1 r_1 - o_1 r_1 2^h$  sets contains exactly  $2^h(r_1 + g_1)$  elements from  $RS_1$ . If  $i = 1$  and  $x = z$  then for  $j$  in the first half of interval  $[n_{x-1} + 1, n_x]$  the  $VS_j^1$

contain  $2^h(r_1 + g_1) - 1$  elements and in the second half they contain  $2^h(r_1 + g_1) + 1$  elements from  $RS_1$ . The remainder of the  $f_1 r_1 - o_1 r_1 2^h - (n_x - n_{x-1})$  sets contain exactly  $2^h(r_1 + g_1)$  elements from  $RS_1$ . If  $i \in [2, x]$  then for  $j$  in the first half of interval  $[n_{i-2} + 1, n_{i-1}]$  the  $VS_j^i$  contain  $2^h(r_i + g_i) - 1$  elements and in the second half they contain  $2^h(r_i + g_i) + 1$  elements from  $RS_i$ . The remainder of the  $f_i r_i - o_i r_i 2^h - (n_{i-1} - n_{i-2})$  sets contain exactly  $2^h(r_i + g_i)$  elements from  $RS_i$ . Therefore for each  $i \in [1, x]$  we have that the number of elements in  $VS_j^i$  which are subsets of  $RS_i$  is exactly the number of elements in  $RS_i$ .

We also have to show that  $f_i r_i - o_i r_i 2^h - (n_{i-1} - n_{i-2}) \geq 0$  and that  $f_1 r_1 - o_1 r_1 2^h - (n_x - n_{x-1}) \geq 0$ . We have for any  $i, k \in [1, x]$  that  $o_i r_i 2^h + (n_k - n_{k-1}) \leq 3(\max_{i \in [1, x]} \{r_i + g_i\})^2 2^h$ . On the other hand we have for any  $i \in [1, x]$  that

$$f_i > \frac{2^{h+1}(\max_{i \in [1, x]} \{r_i + g_i\})^4 - \max_{i \in [1, x]} \{r_i + g_i\}}{\max_{i \in [1, x]} \{r_i + g_i\}}. \text{ Since } \max_{i \in [1, x]} \{r_i + g_i\} \geq 2 \text{ we have that}$$

$$2^{h+1}(\max_{i \in [1, x]} \{r_i + g_i\})^3 > 3(\max_{i \in [1, x]} \{r_i + g_i\})^2 2^h + 1 \text{ and therefore } f_i > o_i r_i 2^h + (n_k - n_{k-1}).$$

There are  $f_i g_i + o_i$  sets such that  $VS_j^i \subset GS_i$ . Of those  $f_i g_i + o_i$  sets there are  $o_i r_i 2^h$  sets that contain  $2^h(r_i + g_i) - 1$  elements from  $GS_i$ , these are the sets  $VS_j^i$  where  $j$  is in the second half of the interval  $[n_{i-1} + 1, n_i]$ . The remainder of the  $f_i g_i + o_i - o_i r_i 2^h$  sets contain exactly  $2^h(r_i + g_i)$  elements. Therefore for each  $i \in [1, x]$  we have that the number of elements in  $VS_j^i$  which are subsets of  $GS_i$  is exactly the number of elements in  $GS_i$ . We have that  $\bigcup_{j \in [1, 2^m]} VS_j^i = GS_i \cup RS_i$  and since  $\bigcup_{vs \in GS_i \cup RS_i} vs = S(a') \cap S(b_i)$  we have that (Req1) is satisfied for  $i \in [1, x]$ .

We now verify the requirement (Req2). For each  $j \in [1, 2^m]$  we have that either for all  $i \in [1, z]$  the sets  $VS_j^i$  contain  $2^h(r_i + g_i)$  elements or there are some two indices  $ip, im$  such that  $|VS_j^{ip}| = 2^h(r_{ip} + g_{ip}) + 1$ ,  $|VS_j^{im}| = 2^h(r_{im} + g_{im}) - 1$  and for all other  $i$ ,  $i \in [1, z] \setminus \{ip, im\}$  we have  $|VS_j^i| = 2^h(r_i + g_i)$ .

It follows that for each  $j$   $\sum_{i \in [1, z]} |VS_j^i| = \sum_{i \in [1, z]} 2^h(r_i + g_i) = 2^{h+d}$ . Since each of the  $2^{h+d}$  intervals has measure  $\lambda(vs) = 2^{-l(w(a'))-d-m-h}$  we have that  $\lambda(\bigcup_{i=1}^z UVS_j^i) = 2^{-m} \lambda(S(a'))$ .

Verifying (Req3) is straightforward from the construction since in each of the sets  $VS_j^i$  we put elements either from set  $RS_i$  or from  $GS_i$ .

We now verify (Req4). Denote

$$t_{i,j} = \frac{\lambda(w) \lambda(UVS_j^i)}{2^{-m}(\lambda(S(a')) + L(a'))(\lambda(S(b_i)) + L(b_i))}$$

Since  $\lambda(UVS_j^i) = |VS_j^i| 2^{-l(w(a'))-d-m-h}$  we have that

$$(2^h(r_i + g_i) - 1) 2^{-l(w(a'))-d-m-h} \leq \lambda(UVS_j^i) \leq (2^h(r_i + g_i) + 1) 2^{-l(w(a'))-d-m-h}$$

From  $\lambda(S(a') \cap S(b_i)) = 2^{-l(w(a'))-d}(r_i + g_i)$  we get

$$t_i - \frac{2^{-l(w(a'))-d-h}\lambda(w)}{(\lambda(S(a')) + L(a'))(\lambda(S(b_i)) + L(b_i))} \leq t_{i,j} \leq t_i + \frac{2^{-l(w(a'))-d-h}\lambda(w)}{(\lambda(S(a')) + L(a'))(\lambda(S(b_i)) + L(b_i))}$$

From (2) we have  $\frac{2^{-l(w(a'))-d}\lambda(w)}{(\lambda(S(a')) + L(a'))(\lambda(S(b_i)) + L(b_i))} \leq 2$ . We have  $t_i - 2^{-h+1} \leq t_{i,j} \leq t_i + 2^{-h+1}$  and therefore  $1 - cs < t_{i,j} < 1 + cs$ .

### 2.5.2

Since the sets of sequences in nodes of the strategies and  $p_n$  are clopen sets there is some integer  $d$  such that there is a set of basic intervals

$$V = \{v : \lambda(v) = 2^{-d}, v \subseteq p_n \vee v \cap p_n = \{\}, \exists u, r, u \in LTN_n^A, r \in LTS_n^B, v \subseteq S(u) \cap S(r)\}$$

We have that  $\bigcup_{v \in V} v = \bigcup_{u \in LTN_n^A} S(u) = \bigcup_{r \in LTS_n^B} S(r)$ . Denote

$$VR = \{v : v \in V, v \subseteq p_n\}, VG = \{v : v \in V, v \cap p_n = \{\}\}$$

we have that  $V = VR \cup VG$ . Index the nodes in  $LTN_n^A$  and  $LTS_n^B$ ,  $LTN_n^A = \{a_1, \dots, a_{za}\}$ ,  $LTS_n^B = \{b_1, \dots, b_{zb}\}$ . Denote

$$t_j^i = \frac{\lambda(w)\lambda(S(a_i) \cap S(b_j))}{(\lambda(S(a_i)) + L(a_i))(\lambda(S(b_j)) + L(b_j))}, a_i \in LTN_n^A, b_j \in LTS_n^B$$

We have that  $1 - cs < t_j^i < 1 + cs$  if  $S(a_i) \cap S(b_j) \neq \{\}$ . Find integer  $h$  such that  $h > 2d$  and

$$\forall i, j, i \in [1, za], j \in [1, zb], 1 - cs < t_j^i(1 - 2^{-h+d+1}), t_j^i(1 + 2^{-h+d+1}) < 1 + cs \quad (10)$$

Denote

$$VRa_i = \{v : v \in VR, v \subseteq S(a_i), a_i \in LTN_n^A\}, ra_i = |VRa_i|$$

$$VGa_i = \{v : v \in VG, v \subseteq S(a_i), a_i \in LTN_n^A\}, ga_i = |VGa_i|$$

Denote

$$VRb_i = \{v : v \in VR, v \subseteq S(b_i), b_i \in LTS_n^B\}, rb_i = |VRb_i|$$

$$VGb_i = \{v : v \in VG, v \subseteq S(b_i), b_i \in LTS_n^B\}, gb_i = |VGb_i|$$

Index the basic intervals in  $VG$ .  $VG = \{vg_1, \dots, vg_z\}$ . Each of the basic intervals in  $VG$  is a union of  $2^h$  basic intervals of measure  $2^{-d-h}$ . Denote

$$VGS = \{v : \lambda(v) = 2^{-d-h}, \exists v' \in VG, v \subset v'\}$$

Index the basic intervals in  $VGS$ ,  $VGS = \{vg_{1,1}, \dots, vg_{z,2^h}\}$  so that for  $vg_{i,j} \in VGS$  there is  $vg_i \in VG$ ,  $vg_{i,j} \subset vg_i$ . Each of the basic intervals in  $VGS$  is a union of  $2^h$  basic intervals of measure  $2^{-d-2h}$ . Denote

$$VGSS = \{v : \lambda(v) = 2^{-d-2h}, \exists v' \in VGS, v \subset v'\}$$

Index the basic intervals in  $VGSS$ ,  $VGSS = \{vg_{1,1,1}, \dots, vg_{z,2^h,2^h}\}$  so that for  $vg_{i,j,k} \in VGSS$  there is  $vg_{i,j} \in VGS$ ,  $vg_{i,j,k} \subset vg_{i,j}$ . For each  $a_i \in LTN_n^A$  we add child nodes such that the leaf-nodes  $a_{i,j}$  have measure  $\lambda(a_{i,j}) = 2^{-d}$ . There are  $2^{d-l(w(a_i))}$  leaf nodes added to each  $a_i$ ,  $\{a_{i,1}, \dots, a_{i,2^{d-l(w(a_i))}}\}$ . For  $j \in [1, ra_i]$  we set  $S(a_{i,j}) = v$ ,  $v \in VRa_i$  and for  $j' \in [1, ra_i]$ ,  $j' \neq j$  we have  $S(a_{i,j}) \neq S(a_{i,j'})$ . For leaf-nodes  $a_{i,ra_i+1}, \dots, a_{i,2^{d-l(w(a_i))}}$  we distribute basic intervals from  $VGS$  among them. Define

$$nexta(j) = \begin{cases} ra_i + 1 & \text{if } j = 2^{d-l(w(a_i))} \\ j + 1 & \text{otherwise} \end{cases}$$

Define

$$nextv(i, j) = \begin{cases} (\min\{o : vg_o \in VG, vg_o \subseteq S(a_i), o > i\}, 1) & \text{if } j = 2^h \\ (i, j + 1) & \text{otherwise} \end{cases}$$

Step0:

For all  $j \in [ra_i + 1, \dots, 2^{d-l(w(a_i))}]$  set  $S(a_{i,j}) := \{\}$   
set  $j := ra_i + 1$ ,  $(w, x) := (\min\{o : vg_o \in VG, vg_o \subseteq S(a_i)\}, 1)$

Step1:

Set  $S(a_{i,j}) := S(a_{i,j}) \cup vg_{w,x}$ ,  $vg_{w,x} \in VGS$

If  $x = 2^h$  and  $w = \max\{o : vg_o \in VG, vg_o \subseteq S(a_i)\}$  finish, otherwise set  $(w, x) := nextv(w, x)$ ,  $j := nexta(j)$  and goto Step1

For each  $b_i \in LTS_n^B$  we add child nodes such that the leaf-nodes  $b_{i,j}$  have measure  $\lambda(b_{i,j}) = 2^{-d}$ . There are  $2^{d-l(w(b_i))}$  leaf nodes added to each  $b_i$ ,  $\{b_{i,1}, \dots, b_{i,2^{d-l(w(b_i))}}\}$ . For  $j \in [1, rb_i]$  we set  $S(b_{i,j}) = v$ ,  $v \in VRb_i$  and for  $j' \in [1, rb_i]$ ,  $j' \neq j$  we have  $S(b_{i,j}) \neq S(b_{i,j'})$ . For leaf-nodes  $b_{i,rb_i+1}, \dots, b_{i,2^{d-l(w(b_i))}}$  we distribute basic intervals from  $VGSS$  among them. Define

$$nextb(j) = \begin{cases} rb_i + 1 & \text{if } j = 2^{d-l(w(b_i))} \\ j + 1 & \text{otherwise} \end{cases}$$

Define

$$nextv(i, j, k) = \begin{cases} (\min\{o : vg_o \in VG, vg_o \subseteq S(b_i), o > i\}, 1, 1) & \text{if } j = 2^h, k = 2^h \\ (i, j + 1, 1) & \text{if } k = 2^h \\ (i, j, k + 1) & \text{otherwise} \end{cases}$$

Step0:

For all  $j \in [rb_i + 1, \dots, 2^{d-l(w(b_i))}]$  set  $S(b_{i,j}) := \{\}$   
set  $j := rb_i + 1$ ,  $(w, x, y) := (\min\{o : vg_o \in VG, vg_o \subseteq S(b_i)\}, 1, 1)$

Step1:

Set  $S(b_{i,j}) := S(b_{i,j}) \cup vg_{w,x,y}, vg_{w,x,y} \in VGSS$

if  $y = 2^h, x = 2^h, w = \max\{o : vg_o \in VG, vg_o \subseteq S(b_i)\}$  finish, otherwise  
 set  $(w, x, y) := nextv(w, x, y), j := nextb(j)$  and goto Step1

For some leaf-nodes  $a_{i,k}, b_{j,l}$  where  $k \in [ra_i + 1, 2^{d-l(w(a_i))}]$ ,  $l \in [rb_j + 1, 2^{d-l(w(b_j))}]$ ,  $S(a_i) \cap S(b_j) \neq \{\}$ ,  $S(a_i) \cap S(b_j) \not\subseteq p_n$  denote

$$E = \{v : v \in VG, v \subset S(a_i) \cap S(b_j)\}$$

$$ES = \{v : v \in VGS, v \subset S(a_{i,k}) \cap S(b_j)\}$$

$$ESS = \{v : v \in VGSS, v \subset S(a_{i,k}) \cap S(b_{j,k})\}$$

We have  $|E| \lfloor \frac{2^h}{ga_i} \rfloor \leq |ES| \leq |E|(\lfloor \frac{2^h}{ga_i} \rfloor + 1)$  and  $|ES| \lfloor \frac{2^h}{gb_j} \rfloor \leq |ESS| \leq |ES|(\lfloor \frac{2^h}{gb_j} \rfloor + 1)$ ,  $\lambda(S(a_i) \cap S(b_j)) = 2^{-d}|E|$  and  $\lambda(S(a_{i,k}) \cap S(b_{j,l})) = 2^{-d-2h}|ESS|$ .  
 From  $\lfloor \frac{2^h}{ga_i} \rfloor = \frac{2^h - oa_i}{ga_i}$ ,  $\lfloor \frac{2^h}{gb_j} \rfloor = \frac{2^h - ob_j}{gb_j}$  and  $oa_i < ga_i < 2^d$ ,  $ob_j < gb_j < 2^d$  we have

$$\frac{\lambda(S(a_i) \cap S(b_j))}{ga_i gb_j} (1 - 2^{d+1-h}) \leq \lambda(S(a_{i,k}) \cap S(b_{j,l})) \leq \frac{\lambda(S(a_i) \cap S(b_j))}{ga_i gb_j} (1 + 2^{d+1-h})$$

From  $\lambda(S(a_{i,k})) = 2^{-d} = \frac{\lambda(S(a_i))}{ra_i + ga_i}$  and (8) we have  $L(a_{i,k}) = \frac{1}{ga_i}(L(a_i) + ra_i \frac{\lambda(S(a_i))}{ra_i + ga_i})$  and  $\lambda(S(a_{i,k})) + L(a_{i,k}) = \frac{1}{ga_i}(\lambda(S(a_i)) + L(a_i))$ . Analogously we have  $\lambda(S(b_{j,l})) + L(b_{j,l}) = \frac{1}{gb_j}(\lambda(S(b_j)) + L(b_j))$

Denote

$$t_{j,l}^{i,k} = \frac{\lambda(w)\lambda(S(a_{i,k}) \cap S(b_{j,l}))}{(\lambda(S(a_{i,k})) + L(a_{i,k}))(\lambda(S(b_{j,l})) + L(b_{j,l}))}$$

from (10) we have  $1 - cs < t_{j,l}^{i,k} < 1 + cs$ .

### 2.5.3

For node  $v \in LTS_i^B$  denote  $g(v) = \frac{\lambda(w)}{m}ms(v)/(L(v) + \lambda(S(v)))$ . We show that

$$L(v) \geq \frac{m}{2c\lambda(w)}(1 - g(v))(\lambda(S(v)) + L(v)) \quad (11)$$

For  $v \in LTS_1^B$  we have  $L(v) = 0$ ,  $ms(v) = m$ ,  $S(v) = w$  and (11) is valid. Assume that for some node  $v' \in LTS_i^B$  (11) is valid. For  $v \in LTS_{i+1}^B$ ,  $v' \prec v$  we have from (7) that

$$ms(v) = (ms(v') - \Delta m)\lambda(S(v))/\lambda(S(v') \setminus p_i)$$

where  $\Delta m = \sum_{a' \in LTN_i^A} d^B(a', v')$ . From (8) we have

$$L(v) + \lambda(S(v)) = (L(v') + S(v'))\lambda(S(v))/\lambda(S(v') \setminus p_i) \quad (12)$$

From (7) we have

$$g(v) = \frac{\lambda(w)}{m}(ms(v') - \triangle m)/(L(v') + \lambda(S(v')))$$

From (8),  $\triangle m \leq 2c\lambda(S(v') \cap p_i)$  and the assumption that (11) is valid for  $v'$  we get that (11) is valid for  $v$ .

Denote the remaining mass for node  $v$  after MAS  $j$  with  $ms_j(v)$ . We have that  $ms_j(v) = ms(v) - \sum_{u \in E_{v,j}} d^B(u, v)$ . Denote  $g_j(v) = \frac{\lambda(w)}{m}ms_j(v)/(L(v) + S(v))$ .

From  $\sum_{u \in E_{v,j}} d^B(u, v) \leq 2c \sum_{u \in E_{v,j}} \lambda(S(u) \cap S(v))$  and (11) we have

$$L(v) + \sum_{u \in E_{v,j}} \lambda(S(u) \cap S(v)) \geq \frac{m}{2c\lambda(w)}(1 - g_j(v))(L(v) + S(v)) \quad (13)$$

If in some iteration  $i$  in some MAS  $j$  the pair of nodes  $(u, v)$  was picked,  $lnp(i, j) = (u, v)$ , then since  $\lambda(S(u) \cap S(v)) \geq \frac{1}{2c}d^A(u, v)$  we have from (2) that  $L(v) + S(v) > \frac{m}{2c(1+cs)}(f - f')$ . From (13) we have that

$$L(v) + \sum_{u' \in E_{v,j-1}} \lambda(S(u') \cap S(v)) > \frac{m^2}{4\lambda(w)c^2(1+cs)}(1 - g_{j-1})(f - f') \quad (14)$$

If for this pair of nodes we have  $d^A(u, v) > 0$ , from definition of MAS we have two cases:

(case1)

$$d^A(u, v) = 2c\lambda(S(u) \cap S(v)), \quad d^B(u, v) = 0$$

We have that  $g_{j-1} \leq f' < 1$  and from (14) we have

$$\lambda(S(u) \cap S(v) + L(v) + \sum_{u' \in E_{v,j-1}} \lambda(S(u') \cap S(v))) > \frac{m^2}{4\lambda(w)c^2(1+cs)}(1 - f')(f - f')$$

since  $\lambda(S(u) \cap S(v)) > 0$ . We have

$$\frac{m^2}{4c^2\lambda(w)(1+cs)}(1 - f')(f - f') > \frac{m^2(1 - cs)}{4\lambda(w)c^2(1+cs)^2}(1 - f')(f - f')$$

since  $0 < cs < 1$ .

(case2)

$$d^A(u, v) + d^B(u, v) = 2c\lambda(S(u) \cap S(v)), \quad d^B(u, v) > 0 \quad (15)$$

We have that  $d^A(u, v) = \frac{m}{\lambda(w)}(f - f')(L(u) + S(u))$ . Since  $f' = g_j$  we have that  $d^B(u, v) = \frac{m}{\lambda(w)}(g_{j-1} - f')(L(v) + S(v))$ . From (2) we have  $(L(u) + S(u))(L(v) + S(v)) > \frac{\lambda(w)}{(1+cs)}\lambda(S(u) \cap S(v))$ . Since (15)

$$\frac{(L(u) + S(u))(L(v) + S(v))}{(f - f')(L(u) + S(u)) + (g_{j-1} - f')(L(v) + S(v))} > \frac{m}{2c(1+cs)}$$

By minimizing the value  $(L(u) + S(u))(L(v) + S(v))$  in previous expression we have

$$(L(u) + S(u))(L(v) + S(v)) > 2\left(\frac{m}{2c(1+cs)}\right)^2(f-f')(g_{j-1}-f')$$

From (3) we have  $\lambda(S(u) \cap S(v)) > \frac{(1-cs)}{\lambda(w)}(L(u) + S(u))(L(v) + S(v))$  and therefore

$$\lambda(S(u) \cap S(v)) > \frac{m^2(1-cs)}{2\lambda(w)c^2(1+cs)^2}(f-f')(g_{j-1}-f')$$

Since  $0 < cs < 1$  we have from (14)

$$\lambda(S(u) \cap S(v)) + L(v) + \sum_{u' \in E_{v,j-1}} \lambda(S(u') \cap S(v)) > \frac{m^2(1-cs)}{4\lambda(w)c^2(1+cs)^2}(1-f')(f-f')$$

#### 2.5.4

For some  $v' \in LTS_i^B$  we have from (6)  $\sum_{v \in LTS_{i+1}^B, v' \prec v} \lambda(S(v)) = \lambda(S(v') \setminus p_i)$ . If

$S(v') \setminus p_i \neq \{\}$  from (8) we have

$$\sum_{v \in LTS_{i+1}^B, v' \prec v} L(v) = L(v') + \lambda(S(v') \cap p_i)$$

If  $S(v') \subseteq p_i$  then there are no nodes  $v \in LTS_{i+1}^B$  such that  $v' \prec v$ . For some  $n' < n$  and  $v' \in LTS_{n'}^B$  we have

$$\begin{aligned} & \sum_{i=n'}^{n-1} \sum_{v \in LTS_i^B, v' \preceq v, S(v) \subseteq p_i} (L(v) + \lambda(S(v))) + \\ & \sum_{v \in LTS_n^B, v' \prec v} (L(v) + \lambda(S(v) \cap p_n)) = L(v') + \sum_{i=n'}^n \lambda(S(v') \cap p_i) \end{aligned}$$

For some  $n$  and some prefix-free set of nodes  $V = \{v : \exists i, i \leq n, v \in LTS_i^B, v' \in V \implies v \not\prec v' \wedge v' \not\prec v\}$  denote  $NV_i = \{v : v \in LTS_i^B, S(v) \subseteq p_i, \forall v' \in V, v' \not\prec v, v \not\prec v'\}$ . From  $L(b) = 0$ ,  $S(b) = w$ ,  $P \subset w$  we have



$$\begin{aligned}
& \sum_{v \in V} (L(v) + \lambda(S(v) \cap p_{\text{indi}(v)})) + \\
& \sum_{v \in V} \sum_{i \in [\text{indi}(v)+1, n]} \lambda(S(v) \cap p_i) + \\
& \sum_{i=1}^{n-1} \sum_{v \in NV_i} (L(v) + \lambda(S(v))) + \\
& \sum_{v \in LTN_n^B, \forall v' \in V, v' \not\prec v} L(v) + \lambda(S(v) \cap p_n) = P_n
\end{aligned}$$

Therefore

$$\sum_{v \in V} (L(v) + \lambda(S(v) \cap p_{\text{indi}(v)})) \leq P_n$$

For some set of nodes  $AV = \{a_1, \dots, a_n\}$ ,  $a_i \prec a_{i+1}$ ,  $a_i \in LTN_i^A$  there is a set of nodes  $BV = \{v : \exists i, \text{indi}(v) = i, S(v) \cap S(a_i) \subseteq p_i\}$ .

For  $a' \in LTN_i^A$ ,  $b' \in LTN_i^B$  we have that if  $S(b') \cap S(a') \subseteq p_i$  then for all  $j > i$ ,  $a'' \in LTN_j^A$ ,  $a' \prec a''$ ,  $b'' \in LTN_j^B$ ,  $b' \prec b''$  we have  $S(a'') \cap S(b'') = \{\}$ . We also have that if  $v, v' \in LTN_i^B$  then  $v \not\prec v' \wedge v' \not\prec v$  and therefore we have  $v, v' \in BV \implies v \not\prec v' \wedge v' \not\prec v$ . Therefore  $\sum_{v \in BV} (L(v) + \lambda(S(v) \cap p_{\text{indi}(v)})) \leq P_n$ .

Since for  $v \in BV$ ,  $a_{\text{indi}(v)} \in AV$  we have

$$\lambda(S(a_{\text{indi}(v)}) \cap S(v)) + \sum_{u' \in E_{v, \text{indj}(a_{\text{indi}(v)}, v)-1}} \lambda(S(u') \cap S(v)) \leq \lambda(S(v) \cap p_{\text{indi}(v)})$$

It follows

$$\sum_{v \in BV} (L(v) + \lambda(S(v) \cap a_{\text{indi}(v)})) + \sum_{u' \in E_{v, \text{indj}(a_{\text{indi}(v)}, v)-1}} \lambda(S(u') \cap S(v)) \leq P_n$$

## 2.5.5

For some leaf-node pair  $(u, v)$  that was picked in MAS  $j$  of iteration  $i$ ,  $\text{lnp}(i, j) = (u, v)$ , denote with  $f(u, v)$  the value

$$f(u, v) = \frac{\lambda(w)}{m} m s_{\text{indj}(u, v)-1}(u) / (\lambda(S(u)) + L(u))$$

and with  $f'(u, v)$  the value

$$f'(u, v) = \frac{\lambda(w)}{m} m s_{\text{indj}(u, v)}(u) / (\lambda(S(u)) + L(u))$$

From (9) we have that

$$\begin{aligned}
& \lambda(S(u) \cap S(v)) + L(v) + \sum_{u' \in E_v, \text{indj}(u,v)=1} \lambda(S(u') \cap S(v)) \\
& > \text{const} \sum_{(u,v) \in F} (1 - f'(u,v))(f(u,v) - f'(u,v))
\end{aligned}$$

We have that  $f'(u,v) \neq f(u,v)$  iff  $d^A(u,v) > 0$ . Denote the remaining mass of a node  $u \in LTN_i^A$  after the last MAS of iteration  $i$  with  $ms'(u)$ . The mass of a node before the first MAS of iteration  $i$  is  $ms(u)$ . For two nodes  $u \in LTN_i^A$ ,  $u' \in LTN_{i+1}^A$ ,  $u \prec u'$  we have from (7) and (12)  $ms(u')/(\lambda(S(u')) + L(u')) = ms'(u)/(\lambda(S(u)) + L(u))$ . Therefore we can order the node pairs in  $F = \{(u_1, v_1), \dots, (u_n, v_n)\}$  so that  $f(u_i, v_i) = f'(u_{i-1}, v_{i-1})$  and  $f(u_1, v_1) = 1$ . From the assumption that the mass of node  $u_n$  after MAS  $t$  of iteration  $r$  is non-positive we have  $f'(u_n, v_n) \leq 0$ . Denote with  $\Delta_i = f(u_i, v_i) - f'(u_i, v_i)$ . We have

$$\sum_{(u,v) \in F} (1 - f'(u,v))(f(u,v) - f'(u,v)) = \sum_{i \in [1, n]} \Delta_i \sum_{j \in [1, i]} \Delta_j$$

Note that

$$2 \sum_{i=1}^n \Delta_i \sum_{j=1}^i \Delta_j = \left( \sum_{i=1}^n \Delta_i \right)^2 + \sum_{i=1}^n \Delta_i^2$$

and since  $\sum_{i=1}^n \Delta_i \geq 1$  we have

$$\sum_{(u,v) \in F} (1 - f'(u,v))(f(u,v) - f'(u,v)) \geq \frac{1}{2}$$

## References

- [1] M. Li, P. Vitanyi An Introduction to Kolmogorov Complexity and Its Applications, second edition Springer, 1997.